U.S. Department of Justice
Office of Justice Programs
Bureau of Justice Statistics

1998 Statistical Tables for the National Crime Victimization Survey

Table of contents

May 2000, NCJ 181585

Survey Methodology for Criminal Victimization in the United States

The survey results contained in this report are based on data gathered from residents living throughout the United States, including persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates, were not included in the scope of this survey. Similarly, U.S. citizens residing abroad and foreign visitors to this country were excluded. With these exceptions, individuals age 12 or older living in units selected for the sample were eligible to be interviewed.

Data Collection

Each housing unit selected for the National Crime Victimization Survey (NCVS) remains in the sample for 3 years, with each of seven interviews taking place at 6-month intervals. An NCVS interviewer's first contact with a housing unit selected for the survey is in person. The interviewer may then conduct subsequent interviews by telephone.

To elicit more accurate reporting of incidents, NCVS uses the self-respondent method which calls for the direct interviewing of each person 12 years or older in the household. An exception is made to use proxy interviewing instead of direct interviewing for the following three cases: 12- and 13-year-old persons when a knowledgeable household member insists they not be interviewed directly, incapacitated persons, and individuals absent from the household during the entire field-interviewing period. In the case of temporarily absent household members and persons who are physically or mentally incapable of granting interviews, interviewers may accept other household members as proxy respondents, and in certain situations non-household members may provide information for incapacitated persons.

Some interviews were conducted using Computer-Assisted Telephone Interviewing (CATI), a data collection mode which involves interviewing from centralized facilities and using a computerized instrument. In the CATI-eligible part of the sample, all interviews are done by telephone whenever possible, except for the first interview, which is primarily conducted in person. The telephone interviews are conducted by the CATI facilities in Hagerstown, Maryland and Tucson, Arizona.

Sample Design and Size

Survey estimates are derived from a stratified, multi-stage cluster sample. The primary sampling units (PSU's) composing the first stage of the sample were counties, groups of counties, or large metropolitan areas. Large PSU's were included in the sample automatically and are considered to be self-representing (SR) since all of them were selected. The remaining PSU's, called non-self-representing (NSR), because only a subset of them was selected, were combined into strata by grouping PSU's with similar geographic and demographic characteristics, as determined by the 1990 Census.

The initial 1990 design consisted of 93 SR PSU's and 152 NSR strata, with one PSU per stratum selected with probability proportionate to population size. A sample reduction was done in October of 1996, reducing the number of NSR PSUs by 42. So, the current NCVS sample consists of only 110 NSR PSUs. The NCVS sample design continues use of both 1980- and 1990-based sample through 1997. Beginning in 1998 only 1990-based sample remains.

In the second stage of sampling, each selected stratification PSU is divided into four nonoverlapping frames (unit, area, permit, and group quarters) from which NCVS independently selects its sample.

From each selected stratification PSU, clusters of approximately four housing units or housing unit equivalents are selected from each frame. For the unit and group quarters frames, addresses come from the 1990 Census. For the permit frame, addresses come from building permit data obtained from building permit offices. This ensures that units built after the 1990 Census are included in the sample. For the area frame, sample blocks come from the 1990 Census files. Then, addresses are listed and sampled in the field.

In order to conduct field interviews, the sample is divided into six groups, or rotations, and each group of households is interviewed once every 6 months over a period of 3 years. The initial interview is used to bound the interviews (bounding establishes a time frame to avoid duplication of crimes on subsequent interviews), but is not used to compute the annual estimates. Each rotation group is further divided into six panels. A different panel of households, corresponding to one sixth of each rotation group, is interviewed each month during the 6-month period. Because the survey is continuous, newly constructed housing units are selected as described, and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every 6 months, replacing a group phased out after being in the sample for 3 years.

Selection of Cases for CATI

Currently, the NCVS sample PSU's fall into three groups of CATI usage: maximum-CATI PSU's, where all the segments in the PSU are CATI-eligible; half-CATI PSU's, where half of the segments in the PSU are randomly designated to be CATI-eligible; and no-CATI PSU's, where none of the segments are CATI-eligible. The level of CATI usage for each PSU was established with concern toward an optimal workload for the field interviewers. In the "half-CATI" PSU's, a random sample of about 50% of the segments in each PSU is taken and designated as CATI-eligible. The sample cases in CATI-eligible segments from the max-CATI and the half-CATI PSU's are interviewed from CATI facilities while the other sample cases are interviewed by the standard NCVS field procedures.

Collection Year Estimates

The data in the surveys were collected during the calendar year being estimated. Because of the retrospective nature of the survey, the estimates include some incidents that actually occurred during the previous year. Analyses comparing the victimization information collected in a calendar year (termed a collection year) to that obtained about victimizations experienced in the same calendar year (called a data year) show only a small difference between the two methods. The differences will be greater during periods of changing crime rates and less during periods of stable rates.

Estimation Procedure

Annual collection year estimates of the levels and rates of victimization are derived by accumulating four quarterly estimates. The weights of all crimes reported during interviews in that year are summed, regardless of when the crime occurred. The base for the collection year rate for personal crime is sum of all person weights. Likewise, the base for the property crime rates is the sum of all household weights.

Month of Interview by Month of Reference

(X's denote months in the 6-month reference period)

Month of interview	Period of reference within bounded period											
	First Quarter			Second Quarter			Third Quarter			Fourth Quarter		
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
January												
February	X											
March	X	X										
April	X	X	X									
May	X	X	X	X								
June	X	X	X	X	X							
July	X	X	X	X	X	X						
August		X	X	X	X	X	X					
September			X	X	X	X	X	X				
October				X	X	X	X	X	X			
November					X	X	X	X	X	X		
December						X	X	X	X	X	X	
January							X	X	X	X	X	X
February								X	X	X	X	X
March									X	X	X	X
April										X	X	X
May											X	X
June												X
July												

The estimation procedure begins with the application of a base weight to the data from each individual interviewed. The base weight is the reciprocal of the probability of each unit's selection for the sample, and provides a rough measure of the population represented by each person in the sample. Next, an adjustment is made to account for households and individuals in occupied units who were selected for the survey but unavailable for interview.

In addition to adjusting for unequal probabilities of selection and observation, the final weight also includes a ratio adjustment to known population totals based on the adjusted counts from the 1990 Census. Specifically, the final person weight is the product of the values of the following six component weights; the final household weight is the product of all components except the within-household non-interview adjustment component detailed below:

Probabilities of selection

- Base weight: The inverse of the sample unit's probability of selection.

- Weighting control factor: adjusts for any subsampling due to unexpected events in the field, such as unusually high growth in new construction, area segments larger than anticipated, and other deviations from the overall stratum sampling rate.

Probabilities of observation (Nonresponse)

- Household non-interview adjustment: adjusts for nonresponse at the household level by inflating the weight assigned to interviewed households so that they represent themselves and non-interviewed households.

- Within-household non-interview adjustment: adjusts for nonresponse at the person level by inflating the weight assigned to the interviewed persons so that they represent themselves and the missed interviews.

Post-stratification ratio adjustment to known population totals

The distribution of the sample population may differ somewhat from that of the total population in terms of age, race, sex, residence, and other characteristics. Because of this, two stages of ratio estimation are employed to bring the two distributions into closer agreement, thereby reducing the variability of the sample estimates.

- First-stage factor: the first stage of ratio estimation is applied only to non-self-representing PSU's. Its purpose is to reduce sampling error caused by selecting one PSU to represent an entire stratum. It adjusts for race and zone of residence differences between the sample non-self-representing PSU's and the population non-self-representing PSU's. (For self-representing PSU's this factor is set to 1).

- Second-stage factor: the second stage of ratio estimation is applied on an individual basis to bring the distribution of individuals in the sample into closer agreement with independent current estimates of the population according to age, sex and race characteristics[1]. This factor is defined for each person to adjust for the difference between weighted counts of persons (using the above five weight components) and independent estimates of the number of persons, within the defined cells. These independent estimates are projections based on the 1990 Census population controls adjusted for the undercount.

For household crimes, the characteristics of the wife in a husband-wife household and the characteristics of the head of household in other types of households are used to determine the ratio adjustment factors. This procedure is considered more precise than simply using the characteristics of the head of household since sample coverage is generally better for females than males.

For estimates involving *incidents* rather than *victimizations*, further adjustments are made to those cases where an incident involved more than one person. These incidents have more than one chance of being included in the sample so each multiple-victimization is reduced by the number of victims. Thus, if two

[1] Armed forced personnel who are eligible to be interviewed are not included in the second-stage ratio estimate and receive a factor of 1.

people are victimized during the same incident, the weight assigned to that incident is the person weight reduced by one-half so that the incident cannot be counted twice. However, the details of the event's outcome as they related to the victim are reflected in the survey results. No adjustment is necessary in estimating data on household crimes because each separate crime is defined as involving only one household.

Series Victimizations

A series victimization is defined as six or more similar but separate crimes which the victim is unable to recall individually or describe in detail to an interviewer. These series crimes have been excluded from the tables because the victims were unable to provide details for each event. Data on series crimes are gathered by the calendar quarter(s) of occurrence, making it possible to match the timeframes used in tabulating the data for non-series crimes. (See Table 110).

The effect of combining series and non-series crimes, counting each of the series crimes as a single victimization based on the details of the most recent incident, was included in the initial release of the 1980 data[2]. The report showed that victimization counts and rates were higher in 1979 and 1980 when the series crimes were added. However, rate changes between these 2 years were basically in the same direction and significantly affected the same crimes as those affected when only non-series crimes were analyzed.

Accuracy of Estimates

The accuracy of an estimate is a measure of its total error, that is, the sum of all the errors affecting the estimate: sampling error as well as nonsampling error.
The sample used for the NCVS is one of a large number of possible samples of equal size that could have been obtained by using the same sample design and selection procedures. Estimates derived from different samples would differ from one another due to sampling variability, or sampling error.

The standard error of a survey estimate is a measure of the variation among that estimate from all possible samples. Therefore, it is a measure of the precision (reliability) with which a particular estimate approximates the average result of all possible samples. The estimate and its associated standard error may be used to construct a confidence interval. A confidence interval is a range of numbers which has a specified probability that the average of all possible samples, which is the true unknown value of interest in an unbiased design, is contained within the interval. About 68% of the time, the survey estimate will differ from the true average by less than one standard error. Only 10% of the time will the difference be more than 1.6 standard errors, and just 1 time in 100 will it be greater than 2.5 standard errors. A 95% confidence interval is the survey estimate plus or minus twice the standard error, thus there is a 95% chance that the result of a complete census would fall within the confidence interval.

In addition to sampling error, the estimates in this report are subject to nonsampling error. While substantial care is taken in the NCVS to reduce the sources of nonsampling error throughout all the survey operations, by means of a quality assurance program, quality controls, operational controls, and error-correcting procedures, an unquantified amount of nonsampling error remains still.

Major sources of nonsampling error are related to the inability of the respondents to recall in detail the crimes which occurred during the 6 months prior to the interview. Research based on interviews of victims obtained from police files indicates that assault is recalled with the least accuracy of any crime measured by the NCVS. This may be related to the tendency of victims to not report crimes committed by offenders who are not strangers, especially if they are relatives. In addition, among certain groups, crimes which contain elements of assault could be a part of everyday life, and are therefore forgotten or not considered important enough to mention to a survey interviewer. These recall problems may result in an understatement of the actual rate of assault.

Another source of nonsampling error is the inability of some respondents to recall the exact month a crime occurred, even though it was placed in the correct reference period. This error source is partially offset by interviewing monthly and using the estimation procedure described earlier. Telescoping is another problem in which incidents that occurred before the reference period are placed within the

---

[2]See *Criminal Victimization in the United States; 1979-80 Changes, 1973-80 Trends*, BJS Technical Report, NCJ-80838, July 1982

period. The effect of telescoping is minimized by using the bounding procedure previously described. The interviewer is provided with a summary of the incidents reported in the preceding interview and, if a similar incident is reported, it can be determined whether or not it is a new one by discussing it with the victim. Events which occurred after the reference period are set aside for inclusion with the data from the following interview.

Other sources of nonsampling error can result from other types of response mistakes, including errors in reporting incidents as crimes, misclassification of crimes, systematic data errors introduced by the interviewer, errors made in coding and processing the data. Quality control and editing procedures were used to minimize the number of errors made by the respondents and the interviewers.

Since field representatives conducting the interviews usually reside in the area in which they interview, the race and ethnicity of the field representatives generally matches that of the local population. Special efforts are made to further match field representatives and the people they interview in areas where English is not commonly spoken. About 90% of all NCVS field representatives are female.

Standard errors measure only those nonsampling errors arising from transient factors affecting individual responses completely at random (simple response variance); they do not reveal any systematic biases in the data. As calculated in the NCVS, the standard errors would partially measure nonsampling error arising from some of the above sources, such as transient memory errors, or accidental errors in recording or coding answers, for example.

Computation and Application of Standard Errors

Deriving standard errors which are applicable to a wide variety of items and which can be prepared at a moderate cost requires a number of approximations. Therefore, three generalized variance function (gvf) constant parameters (identified as "a", "b", and "c" in the following section) were developed for use in calculating standard errors. The parameters provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Direct variances were calculated using the balanced repeated replication (BRR) method. The estimates and their corresponding variances were fit to the standard 3-parameter model to obtain the value of the parameters.

NOTATION

x	=	the estimated number (level) of personal or household victimizations or incidents
y	=	the base; either the total number of persons or households (for victimization rates) or the total of all victimizations (for incident characteristics)
p	=	the estimated proportion, resulting from dividing the number of victimizations into the base. Also, the percentage or rate expressed in decimal form. The percentage is *100p* and the rate per thousand is *1000p*.
s(p)	=	the estimated standard error of *p*
		It follows that:
		s(percentage) = *s(100p) = 100 s(p)*
		s(rate) = *s(1000p) = 1000 s(p)*
a,b,c	=	the generalized variance function parameters (see chart)

1996 Parameter Set	a	b	c
1. Overall Person Crime Estimates	-0.00004325	2096	1.683
2. Person Crime Domain Estimates	-0.00006543	2378	1.883
3. Overall Property Crime Estimates	-0.00009283	1567	1.332
4. Property Crime Domain Estimates	-0.00005524	2281	1.203

1997 Parameter Set	a	b	c
1. Overall Person Crime Estimates	.00017479	2792	1.720
2. Person Crime Domain Estimates	.00016972	2945	2.010
3. Overall Property Crime Estimates	.00008155	1299	2.091
4. Property Crime Domain Estimates	.00023869	2278	1.731

1998 Parameter Set	a	b	c
1. Overall Person Crime Estimates	-0.00007044	2029	3.320
2. Person Crime Domain Estimates	0.00001297	2656	3.390
3. Overall Property Crime Estimates	-0.00002708	1717	1.839
4. Property Crime Domain Estimates	0.00003528	2263	1.835

Parameter set #1 is used for the overall person crime estimates (Table 1). These are the person crime estimates by crime category for the whole population, not disaggregated by any victim, offender, or incident characteristics, nor any variable related to reporting to police.

Parameter set #2 is used for the person crime domain estimates. These are the person crime estimates disaggregated by victim, offender, or incident characteristics, or any variable related to reporting to police.

Parameter set #3 is used for the property crime estimates for the whole population (Table 1). These are the property crime estimates by crime category for the whole population, not disaggregated by any household characteristics, nor any variable related to reporting to police.

Parameter set #4 is used for the property crime domain estimates. These are the property crime estimates disaggregated by household characteristics, or any variable related to reporting to police.

For the statistic from Table 1 that corresponds to the crime category "all crimes" (i.e., person and property crimes together), parameter set #3 should be used. When the person and property estimates are combined (i.e., all crimes) and disaggregated by victim, household, incident characteristics, as well as any variable related to reporting to police, parameter set #4 should be used for the best estimate of the corresponding variance. All of the following examples are based upon 1997 data. These formulas are applied in the same way for other years.

*Formula 1***.** Levels: Standard errors for the estimated *number of victimizations or incidents* may be calculated by using the following formula:

$$s(x) = \sqrt{ax^2 + bx + cx^{3/2}}$$

The following example illustrates
the proper use of Formula 1. Table 1 in 1997 shows 606,660 completed robberies. This estimate and the appropriate parameters are substituted in the formula as follows:

$$s(x) = \sqrt{(.00017479)(606{,}660)^2 + (2792)(606{,}660) + (1.720)(606{,}660)^{3/2}} = 50{,}704$$

Therefore, the 95% confidence interval around the estimated number of robbery victimizations is about equal to 606,660 plus or minus twice (1.96) the standard error, or plus or minus 99,379: an interval of 507,281 to 706,039.

Formula 2. Proportions, Percentages, and Rates: Standard errors for the estimated *victimization rates* or *percentages* are calculated using the following formula:

$$s(p) = \sqrt{\frac{b\ p\ (1.0 - p)}{y} + \frac{c\ p\ (\sqrt{p} - p)}{\sqrt{y}}}$$

The following example demonstrates the use of Formula 2. Table 3 in 1997 shows an estimated robbery rate of 7.4 per 1,000 persons between the ages of 20 and 24, based on a total of 17,648,850 persons in this age range. Substituting the appropriate values into the formula yields:

$$s(p) = \sqrt{\frac{2945\ (.0074)\ (1.0-.0074\ )}{17{,}648{,}850} + \frac{2.010\ (.0074)\ (\sqrt{.0074}-.0074)}{\sqrt{17{,}648{,}850}}}$$

$$= \quad 0.0012 \text{ or } 1.2 \text{ per thousand}$$

Thus, the 95% confidence interval is 7.4 per 1,000 plus or minus 2.4: an interval of 5.0 to 9.8 per 1,000.

*Formula 3***.** Difference in rates or percentages with different bases: The standard error of *a difference between two rates or percentages having different bases* is calculated using the formula:

$$s(p_1 - p_2) = \sqrt{var(p_1) + var(p_2) - 2\ \rho\ s(p_1)s(p_2)}$$

where: *rho* is the year-to-year correlation between p_1 and p_2 (see chart); and $var(p_1)$ and $var(p_2)$ are the square of the standard error of p using Formula 2 for each rate and substituting:

p_1 = first percent or rate (expressed as a proportion in decimal form)

y_1 = base from which first percent or rate was derived

p_2 = second percent or rate (expressed as a proportion in decimal form)

y_2 = base from which second percent or rate was derived

Year-to-Year Correlation Between Estimates

Because of the year-to-year overlap in the sample, the same households and persons contribute to annual estimates for different years. This year-to-year correlation between estimates is measured by rho. In general:

P = 0 when estimates are for the same year

rho ≠ 0 for year-to-year comparisons

When comparing estimates that are 1 year apart, use rho as shown below.

When comparing estimates that are 2 years apart, multiply rho by 1/2

When comparing estimates that are more than 2 years apart, assume rho=0.

Following are NCVS year-to-year correlation values for major crime categories.

TYPE OF CRIME	1996-97 CORRELATION	1995-97 CORRELATION	1996-98 CORRELATION	1997-98 CORRELATION
Total Crimes	0.41	0.14	0.20	0.41
Total Personal Crimes	0.30	0.11	0.15	0.30
Crimes of Violence	0.31	0.11	0.15	0.31
Rape/Sexual Assault	0.04	0.01	0.02	0.04
Robbery	0.04	0.01	0.02	0.04
Assault	0.30	0.11	0.15	0.30
Purse Snatching/Pocket	0.03	0.01	0.01	0.03
Total Property Crimes	0.38	0.13	0.19	0.38
Burglary	0.21	0.07	0.10	0.21
Motor Vehicle Theft	0.08	0.03	0.04	0.08
Theft	0.34	0.12	0.17	0.34

If estimates are uncorrelated, *rho* = 0. Hence, omitting the term containing *rho* in the formula will provide an accurate standard error for the difference between uncorrelated estimates. On the other hand, if the two estimates have a strong positive correlation, omitting the last term will cause overestimation of the true standard error. If the numbers have a strong negative correlation, this will cause underestimation of the actual standard error.

The following example illustrates the use of Formula 3. Table 4 in 1997 lists victimization rate for aggravated assault for males as 10.9 per 1,000 and the rate for females as 6.4 per 1,000. The total number of males in the population is 106,598,660 and the total of females, 113,240,440. Noting that *rho* = 0 because the two estimates are for the same year and placing the appropriate values in the formulas yields:

$$var(p_1) = \frac{2945\ (.0109)\ (1.0-.0109)}{106{,}598{,}660} + \frac{2.010\ (.0109)\ (\sqrt{.0109}-.0109)}{\sqrt{106{,}598{,}660}} = 0.000000496$$

$$var(p_2) = \frac{2945\ (.0064)\ (1.0-.0064)}{113{,}240{,}440} + \frac{2.010\ (.0064)\ (\sqrt{.0064}-.0064)}{\sqrt{113{,}240{,}440}} = 0.000000254$$

Standard error of the difference =

$$\sqrt{0.000000496 + 0.000000254} = .00087 \text{ or } .87 \text{ per thousand}$$

The 95% confidence interval around the difference of 4.5 per thousand is approximately the difference plus or minus 1.7 per thousand (a difference between 2.8 and 6.2 per thousand).

The ratio of a difference to the standard error of the difference is the "z score", which is associated with a given statistical level of significance. For example, a ratio with an absolute value of 2.0 (1.96, to be exact) or greater indicates that the difference is significant at the 95% confidence level (or greater); a ratio with an absolute value between 1.6 and 2.0 indicates the difference is significant at a confidence level between 90% and 95%; a ratio with an absolute value less than 1.6 denotes a confidence level less than 90%. In the previous example, the ratio of the difference (.0045) to the standard error (.00087) is 5.17. Thus the aggravated assault rate for males and females was significantly different at a confidence level exceeding 95%.

Formula 4. Differences in percentages of a single response variable with the same base: The standard error of *the difference between two percentages derived from a single response variable with the same base* is calculated using the formula:

$$s(p_1 - p_2) = \sqrt{var(p_1) + var(p_2) - 2\ \rho\ s(p_1)s(p_2)}$$

where

$$\rho = -\sqrt{\frac{p_1 p_2}{q_1 q_2}}$$

$q_1 = 1 - p_1$
$q_2 = 1 - p_2$

All other terms are as defined in Formula 3, except that y_1 and y_2 are the same common base, y.

The following example, which uses Table 43 for 1997, illustrates the use of Formula 4. The proportion of single-offender violent crime victimizations involving relatives was 10.8% and the proportion involving acquaintances (well-known or casual) was 45.2%, out of a total of 6,737,250 single-offender violent crime victimizations. Substituting the appropriate values into the formula gives:

$$var(p_1) = \frac{2945\ (.108)\ (1.0 - .108)}{6{,}737{,}660} + \frac{2.010\ (.108)\ (\sqrt{.108} - .108)}{\sqrt{6{,}737{,}660}} = 0.000061$$

$$var(p_2) = \frac{2945\ (.452)\ (1.0 - .452)}{6{,}737{,}660} + \frac{2.010\ (.452)\ (\sqrt{.452} - .452)}{\sqrt{6{,}737{,}660}} = 0.000185$$

Standard error of the difference

$$\sqrt{0.000061 + 0.000185 + 2\sqrt{0.09987}(0.008)(0.01361)} = 0.018 \text{ or } 1.8 \text{ percent}$$

The confidence interval around the difference at one standard error is from -32.6% to -36.2% (-34.4% plus or minus 1.8%). The ratio of the difference (-0.344) to its standard error (0.018) is -19.11. Since 19.11 is greater than 2.0, the difference between these two percentages is statistically significant at a confidence level exceeding 95%.